



Delusions as harmful malfunctioning beliefs[☆]



Kengo Miyazono

University of Birmingham, 144 European Research Institute Building, Edgbaston, Birmingham B15 2TT, United Kingdom

ARTICLE INFO

Article history:

Received 1 June 2014

Revised 24 September 2014

Accepted 16 October 2014

Available online 18 November 2014

Keywords:

Delusion

Schizophrenia

Malfunction

Harmful dysfunction analysis of disorder

Teleosemantics

ABSTRACT

Delusional beliefs are typically pathological. Being pathological is clearly distinguished from being false or being irrational. Anna might falsely believe that his husband is having an affair but it might just be a simple mistake. Again, Sam might irrationally believe, without good evidence, that he is smarter than his colleagues, but it might just be a healthy self-deceptive belief. On the other hand, when a patient with brain damage caused by a car accident believes that his father was replaced by an imposter or another patient with schizophrenia believes that “The Organization” painted the shops on a street in red and green to convey a message, these beliefs are not merely false or irrational. They are pathological. What makes delusions pathological? This paper explores the negative features because of which delusional beliefs are pathological. First, I critically examine the proposals according to which delusional beliefs are pathological because of (1) their strangeness, (2) their extreme irrationality, (3) their resistance to folk psychological explanations or (4) impaired responsibility-grounding capacities of people with them. I present some counterexamples as well as theoretical problems for these proposals. Then, I argue, following Wakefield’s harmful dysfunction analysis of disorder, that delusional beliefs are pathological because they involve some sorts of harmful malfunctions. In other words, they have a significant negative impact on wellbeing (=harmful) and, in addition, some psychological mechanisms, directly or indirectly related to them, fail to perform the jobs for which they were selected in the past (=malfunctioning). An objection to the proposal is that delusional beliefs might not involve any malfunctions. For example, they might be playing psychological defence functions properly. Another objection is that a harmful malfunction is not sufficient for something to be pathological. For example, false beliefs might involve some malfunctions according to teleosemantics, a popular naturalist account of mental content, but harmful false beliefs do not have to be pathological. I examine those objections in detail and show that they should be rejected after all.

© 2014 Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/3.0/>).

1. Introduction

Delusional beliefs are typically pathological.¹ Being pathological is clearly distinguished from being false or being irrational. Anna might falsely believe that her husband is having an affair but it might just be a simple mistake. Again, Sam might irrationally believe, without good evidence, that he is smarter than his colleagues, but it might just be a healthy self-deception. On the

[☆] This article is part of a special issue of this journal on Imperfect Cognitions.

E-mail address: k.miyazono@bham.ac.uk

¹ However, I do not assume that all delusional beliefs are pathological (see below). In this paper, I only discuss typical delusional beliefs that are pathological. All my references to delusional beliefs without any specifications are going to be about pathological ones.

other hand, when DS, a patient with brain damage caused by a car accident, believes that his father was replaced by an imposter (Hirstein & Ramachandran, 1997) or Peter with schizophrenia, believes that “The Organization” painted the shops on a street in red and green to convey a message (Chadwick, 2001), these beliefs are not merely false or irrational. They are pathological.

What makes delusional beliefs pathological? This paper explores the negative features because of which delusional beliefs are pathological. In Section 2, I critically examine the proposals according to which delusions are pathological because of (1) their strangeness, (2) their irrationality, (3) their resistance to folk psychological explanations or (4) impaired responsibility-grounding capacities of people with them. I present some counterexamples as well as theoretical problems for these proposals. In Section 3, I argue, following Wakefield’s harmful dysfunction analysis of disorder, that delusional beliefs are pathological because they involve some sorts of harmful malfunctions. In other words, they have a significant negative impact on wellbeing (=harmful) and, in addition, some psychological mechanisms, directly or indirectly related to them, fail to perform the jobs for which they were selected in the past (=malfunctioning). An objection to the proposal is that delusional beliefs might not involve any malfunctions. For example, they might be playing psychological defence functions properly. Another objection is that a harmful malfunction is not sufficient for something to be pathological. For example, false beliefs might involve some malfunctions according to teleosemantics, a popular naturalist account of mental content, but harmful false beliefs do not have to be pathological. I examine those objections in detail in Section 4 and show that they should be rejected after all.

The central question of this paper is about what makes delusional beliefs pathological. Before starting, I have several remarks on the idea that delusional beliefs are pathological.

First, when I use the term “pathological” in talking about mental states, I refer to the property of the mental states in virtue of which they constitute, together with other symptoms, mental disorders. Delusional beliefs, together with other positive and negative symptoms, constitute schizophrenia, for example.

Second, the idea that a belief is pathological is different from the idea that it is delusional. Unfortunately, there is no uncontroversial definition of delusionality. According to DSM-5, a delusion is “a false belief based on incorrect inference about external reality that is held despite what almost everyone else believes and despite what constitutes incontrovertible and obvious proof or evidence to the contrary” (American Psychiatric Association, 2013, p. 819). This definition is, however, very controversial. Delusions might be accidentally true. Some delusions are not about external reality but rather about internal mental states. Some delusions might not be based on inference of any sort, and so on. In this paper, I simply stipulate that a belief is delusional if sufficient psychiatrists regard it as delusional.

Third, I assume that delusionality and pathology can come apart, at least, in principle. First, some delusional beliefs might not be pathological. For example, it is often argued that healthy individuals can have delusional beliefs or delusion-like ideas.² A person without any psychiatric diagnosis might have a paranoia belief that his colleagues are trying prevent him from being promoted. The belief is delusional (i.e. regarded as delusional by psychiatrists) but not pathological (i.e. does not constitute a mental disorder). Again, some pathological beliefs might not be delusional. For example, some instances of confabulations or obsessive thoughts might involve non-delusional pathological beliefs. It is conceivable that a person has obsessive thoughts about being contaminated by germs, but the thoughts are not delusional because he perfectly recognizes their implausibility. The thoughts in such a case are pathological (i.e. constitute a mental disorder, such as OCD) but not delusional (i.e. not regarded as delusional by psychiatrists).³

2. Some unsuccessful answers

(1) *Strangeness*: Anna’s belief that his husband is having an affair is false but it is not very strange. Many married women can have the same belief at some point. DS’s belief that his father was replaced by an imposter, on the other hand, is not only false but also strange.⁴ The same thing is true about Peter’s belief that The Organization painted the shops to convey a message. This observation motivates the first proposal, according to which delusions are pathological because their strange content. In other words, the pathology of delusion comes from the abnormality of the content.

A problem of this proposal is that it is not obvious that all delusions are significantly stranger than healthy beliefs. Peter’s belief is certainly strange. But, there are some healthy beliefs that are as strange as his. For example, Murphy (2013) discusses a community in Sudan where it is believed that ebony trees provide important social information. The belief about ebony trees is culturally normal and hence not pathological. Nonetheless, it seems to be as strange as Peter’s delusional belief. One might think, however, that this problem can be solved by introducing a culture-relative notion of strangeness. The idea, for example, is that the belief about ebony trees is not strange relative to the culture in the community. Peter’s belief, on the other hand, is strange relative to the culture in the western, modern community to which he belongs. But, this

² See (Freeman, 2006) for an overview.

³ One might think, however, that the person in this case does not believe the contamination and, thus, this is not an example of pathological beliefs that are not delusional. This is a possible interpretation, but it might not be the only one. It is not utterly implausible to think that the person believes the contamination and this belief explains his non-verbal behavior such as his washing hands repeatedly. See (Bortolotti, 2010) for detailed discussions about the relationship between delusions and obsessive thoughts.

⁴ I am talking about the strangeness of content here. We might also talk about the strangeness of belief-forming processes; Anna’s belief might be formed in a normal way on the basis of sufficient evidence, but Peter’s belief is probably not. This type of strangeness might be regarded as a kind of epistemic rationality, which I will discuss below.

response does not solve all the problems, because pathological delusions and healthy beliefs with similar content can exist in the same cultural contexts. For example, it can be difficult to distinguish Anna's belief from the delusional jealousy in Othello syndrome by content alone. Presumably, the main difference between healthy beliefs about the partner's infidelity and delusional jealousy is not about the content but rather about the sensitivity to the cues. As Easton, Schipper, and Shackelford noted, delusional jealousy "can be thought of as hypersensitive jealousy, as these individuals experience jealous reactions at a much lower threshold than normal individuals" (Easton, Schipper, & Shackelford, 2007, p. 399).

Another problem is that being strange is not sufficient for a belief to be pathological. Philosophers seriously believe very strange things. But, typically, these philosophical beliefs are not the expressions of a mental disorder, but rather of remarkable insights and argumentative skills. For example, there are some philosophers who seriously believe that every single object in the universe is conscious (panpsychism), that, for any objects, however arbitrary they are chosen, there is a further object that is composed by them (unrestricted composition), that there are facts about the boundaries of a vague predicate which we can never discover (epistemicism about vague predicates), and so on.

(2) *Irrationality*: Sam's belief that he is smarter than his colleagues is irrational but, presumably, DS's belief that his father was replaced by an imposter and Peter's belief that The Organization painted the shops to convey a message are more irrational. Maybe, they are too irrational. According to the second proposal, delusion is pathological because of their extreme irrationality.⁵

However, it is not obvious that delusional beliefs are extremely irrational. According to empiricist accounts of delusion formation, which is very influential recently, delusions are formed in response to abnormal experience. Given the fact that the abnormal experience can be understood as a kind of evidence for delusional beliefs, it is not obvious at all that, according to empiricism, delusions are extremely irrational. Indeed, a number of empiricist researchers support the view that a delusion is a reasonable response to abnormal experience. Maher famously argued that delusions "are derived by cognitive activity that is essentially indistinguishable from that employed by non-patients, by scientists, and by people generally" (Maher, 1974, p. 103). Coltheart, Menzies, and Sutton (2010) support this claim and argue that it is perfectly Bayesian rational for a Capgras patient to adopt the imposter hypothesis rather than the competing, realistic hypotheses given the abnormal data. (They do not use the term "experience" because they think that the "data" are not consciously accessible.) The imposter hypothesis actually gets a higher posterior probability than the competing hypotheses. Similarly, Corlett and colleagues argue that it is hard for a Capgras patient to avoid the imposter hypothesis given the abnormal experience they have; "the phenomenology of the percepts are such that bizarre beliefs are inevitable; surprising experiences demand surprising explanations" (Corlett, Taylor, Wang, Fletcher, & Krystal, 2010, p. 360).

Still, the view that a delusion is a rational response to abnormal experience is controversial. Stone and Young (1997), for instance, argue that delusional beliefs are produced not only by abnormal experience but also by the irrational reasoning with the bias toward observational adequacy; people with delusion irrationally put more emphasis on incorporating new observations into belief system (observational adequacy) than keeping existing beliefs as long as possible (doxastic conservatism). McKay (2012) follows this suggestion and argues, in response to Coltheart et al. (2010) that delusional beliefs are produced through the Bayesian-irrational reasoning process with the bias of discounting prior probabilities of the hypotheses; people with delusion irrationally put more emphasis on likelihoods (which summarize how nicely hypotheses explain the observation) than prior probabilities (which summarize how probable the hypotheses are prior to the observation). But, even if one of those views is correct, it is still not obvious that delusional beliefs are extremely irrational. Similar irrational biases might be found in healthy beliefs as well. For example, the famous study by Kahneman and Tversky on the base-rate neglect (1973) shows that normal people have the strong tendency to neglect the base-rate information that is relevant for given hypotheses. Here, the base-rate information gives the prior probabilities of the hypotheses at issue. Thus, the tendency to neglect the base-rate information can be understood as the tendency to neglect prior probabilities.

One of the basic principles of statistical prediction is that prior probability, which summarizes what we knew about the problem before receiving independent specific evidence, remains relevant even after such evidence is obtained. Bayes' rule translates this qualitative principle into a multiplicative relation between prior odds and the likelihood ratio. Our subjects, however, fail to integrate prior probability with specific evidence. [...] The failure to appreciate the relevance of prior probability in the presence of specific evidence is perhaps one of the most significant departures of intuition from the normative theory of prediction (Kahneman & Tversky, 1973, p. 243).

Experimental studies revealed some "biases" in judgment processes of people with delusion. However, the studies do not necessarily support the idea that delusions are very irrational. The term "bias" can be used in, at least, two different ways. It might refer to the deviation from the norm of (logico-mathematical) rationality. When we say that the tendency to neglect the base-rate information is "biased", the term is used in this sense. It deviates from the Bayesian norm of rationality. Alternatively, the term might refer to the deviation from the performance of normal people. In the context of the delusion research, the term "bias" is often used in the second way. So, it is perfectly possible that the "biased" performance of people with delusion does not actually deviate from the norm of rationality. This possibility is nicely illustrated by the well-known study of the "jumping-to-conclusion bias" (Huq, Garety, & Hemsley, 1988). It was found in the study that people with

⁵ Here, I am talking about epistemic irrationality, which is determined by the relationship between beliefs and available evidence. See (Bortolotti, 2010) for the discussion of other sorts of irrationality that can be attributed to delusional beliefs.

delusion have the tendency to “jump to conclusion”; they require less evidence before coming to conclusions than people in control groups (healthy people and non-delusional people with schizophrenia). At the same time, though, it was also found that the performance of people with delusion is more rational from a mathematical point of view than that of people in control groups. People with delusion reach the conclusion when the probability of the hypothesis is reasonably high, while people in control groups do not reach the conclusion until the probability is unnecessarily high. So, Huq et al. wrote; “[i]t may be argued that the deluded sample reached a decision at an objectively ‘rational’ point. It may further be argued that the two control groups were somewhat over cautious” (Huq et al., 1988, p. 809).

(3) *Understandability*: Sam’s belief that he is smarter than his colleagues is irrational, but it is “understandable” in the sense that we can give a simple folk psychological account of it. Sam comes to believe it because he wants it to be the case that he is smarter than his colleagues. In other words, his belief is driven by the desire to be smarter than the colleagues. On the other hand, DS’s belief that his father is replaced by an imposter and Peter’s belief that The Organization painted the shops to convey a message are not “understandable” in this way. There are no easy folk psychological explanations of those delusional beliefs. According to the third proposal, delusions are pathological because of the “ununderstandability” or the resistance to folk psychological explanations.

But, this view is not fully satisfactory. First, it is not clear that the all delusions resist folk psychological explanations. Sam’s belief is “understandable” because we can identify the motivational factors that play crucial roles in the formation of the belief. But, then, when motivational factors play crucial roles for some delusions, at least those delusions are “understandable” by identifying those motivational factors.⁶ Butler (2000) reported the case of B.X. who had the delusion about the fidelity of his former romantic partner (reverse Othello syndrome). B.X. was a gifted musician who had been left quadriplegic following a car accident. One year after his injury, he developed delusional beliefs about the continuing fidelity of his former romantic partner, N., who had in fact severed all contact with him soon after the accident. A pretty straightforward folk psychological explanation of this case would be that B.X. formed his delusional beliefs because he desperately wanted it to be the case that N. still loved him. In other words, his delusional beliefs were driven by the desire that N. still loved him.

Second, resisting folk psychological explanation does not seem to be sufficient for beliefs, or mental states in general, to be pathological. The so-called “twisted self-deception” is a good example. Twisted self-deceptive beliefs are, roughly speaking, the unwelcome irrational beliefs. For example, if it turns out that Anna’s belief about the husband’s affair is not supported by the available evidence at all, it is a twisted self-deceptive belief. Presumably, there is no straightforward folk psychological explanation of twisted self-deceptive beliefs. Mele (1999) provides an influential account according to which twisted self-deceptive beliefs are produced by the people’s tendency to avoid costly errors. If, on one hand, Anna falsely believes that her husband is having an affair, then the falsity is not very costly. For example, it just annoys the husband. If, on the other hand, she falsely believes that the husband is not having an affair, then the falsity is very costly. The relationship is seriously threatened in that case. This account is not purely folk psychological. The idea that people choose their beliefs so that they can avoid costly errors does not seem to be a part of folk psychology. Indeed, the idea comes from the scientific, not folk, psychological theory by Friedrich (1993).

Another example comes from Gendler (2008). Many people experience an extreme fear when they walk on the horse-shoe-shaped transparent walkway on the 4000 feet above the floor of the Grand Canyon. There is nothing abnormal or pathological in this experience. As Gendler noted, “the basic phenomenon—that stepping onto a high transparent safe surface can induce feelings of vertigo—is both familiar and unmysterious” (Gendler, 2008, p. 635). The people who walk on the walkway seem to believe that the walkway is safe. But, then, why do they feel the extreme fear? The fear seems to be ungrounded if they seriously believe that the walkway is safe. Are they, then, somewhat sceptical about the safety? But, in that case, we cannot explain the fact they step onto the walkway in the first place. Presumably, there is no easy folk psychological explanation of the fear. Gendler argues that they feel the extreme fear because they *alieve* that the walkway is not safe, although they believe that it is safe. Alief is, roughly speaking, “a mental state with associatively linked content that is representational, affective and behavioral, and that is activated—consciously or nonconsciously—by features of the subject’s internal or ambient environment” (Gendler, 2008, p. 645). Although alief sounds like another folk psychological state, Gendler’s account of the case is not purely folk psychological. After all, alief is not a part of the conceptual repertoire of folk psychology. Presumably, it is best understood as an extended folk psychological account.⁷

(4) *Responsibility*: Suppose that Anna, on the basis of her belief, acts violently to the woman who is mistakenly regarded as the affair partner. The falsity of the belief does not change the fact that Anna is responsible for what she does. Anna is clearly responsible for her violence. On the other hand, if DS, because of his delusional belief, had acted violently to his father, he would not have been fully responsible for the violence. People with Capgras delusion sometimes act violently on the basis of

⁶ See (Bortolotti & Mameli, 2012) for a similar claim.

⁷ Murphy offers a broader conception of folk theory of mind which includes not just folk psychology in the narrow sense, but also “a much richer body of beliefs and expectations about the role of hot cognition and personal interests in fixing belief” as well as “the role of culture in shaping people’s assumptions about what counts as legitimate evidence.” (Murphy, 2012, p. 22) Does this broad notion of folk theory of mind help the current proposal? Probably it does not. Murphy argues that a belief is regarded as delusional when we are not able to provide the explanation of the belief with the broad folk theory. This is certainly an interesting proposal about the delusionality of delusional beliefs. However, it is far from obvious that the broad conception of folk theory is useful in accounting not only for the delusionality but also the pathology of delusional beliefs. As I already noted, the delusional beliefs in healthy individuals (e.g. paranoid beliefs without any psychiatric conditions) might be the instances of non-pathological delusional beliefs. Murphy might argue that both pathological and healthy delusional beliefs resist the explanation with the broad folk theory, and this explains the fact that they are delusional. But, then, resisting the broad folk theoretical account doesn’t explain the asymmetry in terms of pathology between pathological and healthy delusional beliefs.

their delusional beliefs. In a tragic case in Louisiana in 2011, for instance, a father beheaded his disabled son because he believed that the son was had been replaced by a CPR dummy. Following the testimony by forensic psychiatrists, he was ruled not guilty by reason of insanity. The forth proposal is that delusions are pathological because of responsibility-grounding capacities, such as decision-making capacity or autonomous agency, are significantly impaired in people with delusion.

It is, however, not obvious that responsibility-grounding capacities are always impaired in people with delusion. Certainly, it is very likely that belief-forming or belief-checking capacities are impaired somehow in them. But, the impairment in belief-forming/checking capacities might be dissociated from the impairment in responsibility-grounding capacities.

A possible response to this challenge is that belief-forming/checking capacities are relevant to the attribution of responsibility. In other words, responsibility-grounding capacities include belief-forming/checking capacities. For example, according to M'Naghten rules, a person is not responsible for what he does when he is ignorant of "the nature and quality of the act he was doing." Presumably, people with delusion are ignorant of the nature and quality of their acts because of impaired belief-forming/checking capacities and, hence, they are not responsible for the acts. However, this response is not very convincing. Certainly, there is a sense in which a violent Capgras patient is ignorant of the nature and quality of his act; he thinks that he is attacking the imposter, which is false. But, this is also true about Anna; she thinks that she is attacking the affair partner, which is false. So, if "being ignorant" means "having the false belief" or "not having the true belief", then Anna is as ignorant as violent Capgras patients. The phrase might be interpreted in different ways, but it is not clear that there is an interpretation according to which violent Capgras patients are significantly more ignorant than Anna.

Another response is that delusion is just a tip of iceberg. People with delusion often have other kinds of abnormalities at the same time and these abnormalities impair responsibility-grounding capacities. For instance, delusions in the context of schizophrenia are accompanied by other positive and negative symptoms that directly or indirectly impair responsibility-grounding capacities. However, we cannot assume *a priori* that responsibility-grounding capacities are compromised in people with schizophrenia in general. There might be serious individual differences about the quality of responsibility-grounding capacities of people with schizophrenia given the fact that schizophrenia is an extremely heterogeneous condition. Maybe, the fact that one gets the diagnosis of schizophrenia in itself does not tell us much about his responsibility-grounding capacities. As Bortolotti, Broome, and Mamelì (2013) pointed out, "[t]he assumption that people who have psychotic symptoms or have received a diagnosis of schizophrenia lack responsibility or have reduced responsibility for action is especially problematic, as the behavior of two people with psychosis or schizophrenia can differ almost entirely. Some people with schizophrenia are able to function well, cognitively and socially, and to control their delusions to some extent."

3. HDA and delusion

I believe that the main problem of the previous proposals is that they are detached from the considerations about what, in general, makes a condition pathological. In order to explain why a condition X is pathological, first, we need to have a general account of the features that make conditions pathological and, then, show that X has the features as well. But, the previous proposals skip the first step and simply point out some remarkable negative features of delusion. This invites all sorts of counterexamples and difficulties.

Wakefield (1992a, 1992b) presented a general account of disorder, which is very influential and, in my view, more plausible than its rivals. It is often called "Harmful Dysfunction Analysis of Disorder." I will call it "HDA." According to HDA, disorders are "harmful malfunctions" or "harmful dysfunctions." Being "harmful" means having a negative impact on wellbeing. The harmfulness condition is important because "disorder is in certain respects a practical concept that is supposed to pick out only conditions that are undesirable and grounds for social concern" (Wakefield 1992b, p. 237). A "malfunction" is, roughly speaking, the failure to perform an etiological function,⁸ where the etiological function of something is the performance for which it was selected in the past.⁹ For instance, a heart malfunctions when it fails to pump blood, a kidney malfunctions when it fails to filter metabolic wastes from blood, a corpus callosum malfunctions when it fails to facilitate interhemispheric communications, and so on.

My proposal, which relies on HDA, is that delusions are pathological because (1) they are harmful and (2) they involve some etiological malfunctions directly or indirectly. I will call (1), and (2) "the harmfulness thesis" and "the malfunction thesis" respectively.

There are many objections to HDA. It should be noted, however, that most (if not all) objections are aiming at refuting the necessity claim of HDA, namely, the claim that a harmful etiological malfunction is necessary for a disorder. This means that those objections are not very serious for my purpose. What is crucial for my account is that a harmful etiological malfunction is sufficient for a disorder. We can say that delusions are pathological because they involve harmful etiological malfunction as long as a harmful etiological malfunction is sufficient for a disorder. For example, Tengel (2001) argues that viral infection is a counterexample to HDA. Viral infection is a disorder but its symptoms (e.g. fever, cough, sneezing) are very often biological defences, not malfunctions. Again, Murphy and Woolfolk (2000) argue that appendicitis is another counterexample. Appendicitis is clearly a disorder, but vestigial organs such as appendix cannot fail to perform its etiological functions

⁸ Strictly speaking, this is wrong. I will come back to this in Section 5.

⁹ There is, strictly speaking, a kind of etiological function that does not require past selection but only some past contribution to fitness. It is often called "weak" etiological function (Buller, 1998). The distinction between "strong" and "weak" etiological functions is theoretically important, but not crucially relevant to the following discussions.

because they do not them in the first place.¹⁰ Wakefield argues, in response to Tengland, that viral infection involves etiological malfunctions at the level of cell (Wakefield, 2011) and, in response to Murphy and Woolfolk, that appendicitis involves etiological malfunctions at the level of tissue (Wakefield, 2000). In my opinion, Wakefield's responses are convincing. But, even if they are not, those counterexamples are not very serious for my account, because they are the counterexamples to the necessity of a harmful etiological malfunction for a disorder.

3.1. The harmfulness thesis

Delusions seem to be harmful in all sorts of ways. Delusions do not only cause psychological stress and anxiety but also harmful consequences in the life of the people with them, such as losing a job or failing to maintain relationship with other people. McKay and colleagues propose the following definition of delusions, which explicitly mentions the harmfulness of delusions; "A person is deluded when they have come to hold a particular belief with a degree of firmness that is both utterly unwarranted by the evidence at hand, and that *jeopardises their day-to-day functioning* [emphasis added]" (McKay, Langdon, & Coltheart, 2005, p. 315). Delusions can be especially harmful when people act on them. There are many reported cases where people perform harmful actions on the basis of their delusions. The case in Louisiana that I already mentioned, for example, a father killed his disabled son on the basis of his Capgras delusion, which is not only harmful to the son (being killed), but also to the father (killing his own son without knowing).^{11,12}

Some clarifications are in order.

First, the "harm" at issue does not have to be the harm to people with delusion themselves. It might be the harm to people around them, such as family members, friends, colleagues, or neighbours. It is conceivable that some people with delusions are happy because of them at some point. For example, the grandiose delusion about special abilities given by God might make some people happy at some point. Still, the delusion might be harmful overall because of the serious troubles it causes for other people.

Second, the harmfulness thesis is not about hedonistic pains or pleasures. The life on the experience machine (the machine that produce the perfect illusion of leading whatever kind of life one desires, while one floats about in a tank) is extremely pleasurable from the hedonistic point of view. But, typically, we do not find the life very attractive. Nozick argues that the life is not attractive because "we want to be a certain way, to be a certain sort of person. Someone floating in a tank is an indeterminate blob. There is no answer to the question of what a person is like who has long been in the tank. Is he courageous, kind, intelligent, witty loving? It's not merely that it's difficult to tell; there's no way he is" (Nozick, 1974, p. 43). Similarly, it is conceivable that the life with the grandiose delusion about special abilities is really pleasurable from the hedonistic point of view. But, again, it does not mean that the life is really attractive. It is not attractive because, as a matter of fact, the person with the delusion fails to be the person he really wants to be. He wants to have some special abilities that distinguish him from others. However, he does not have such abilities. His real life might not be less miserable the one in the tank.

Third, there is an implicit *ceteris paribus* clause in the harmfulness thesis. Delusions are *ceteris paribus* harmful. For example, we can imagine the case where Peter avoided a fatal plane crash because he had changed the reservation due to his delusional belief that The Organization wants him to do so. In this case, his delusional belief is beneficial rather than harmful. Actually, the same thing is true about disorders in general. For example, we can easily imagine the case where Peter avoided the plane crash because he had changed his reservation due to his flu. But, of course, the case does not show that flu is not harmful. The claim that flu is harmful has an implicit *ceteris paribus* clause and these tricky cases do not contradict the *ceteris paribus* claim.

3.2. The malfunctioning thesis

According to the malfunctioning thesis, delusions involve some etiological malfunctions. What kinds of malfunctions are they exactly? A fully satisfactory answer to this question is not available yet because the process of delusion formation and maintenance has not been fully understood. Still, the current understanding of the process is informative enough to help us to identify good candidates.¹³

First, according to the empiricist theories, delusions are formed in response to abnormal experience. In this view, there might be some etiological malfunctions in the mechanisms that are responsible for the abnormal experience. For example, Ellis and Young (1990) argue that Capgras delusions are formed in response to the abnormal experience of seeing familiar

¹⁰ The objection presupposes "modern" etiological function (Godfrey-Smith, 1994), which is determined by relatively recent selection for the maintenance of traits.

¹¹ It is, however, not the case that people with delusions always act on the basis of their delusions. Indeed, one of the most surprising facts about delusions is that people with delusions sometimes act inconsistently with the delusions to which they are seriously committed. See (Bortolotti, 2010) for more discussions.

¹² See (Bourget & Whitehurst, 2004) for an overview of the studies of the link between Capgras delusions and violent behavior.

¹³ The following discussion assumes that the relevant malfunctions occur at the level of cognitive architecture. But, this assumption is, strictly speaking, not necessary. For instance, they can be at the level of neurophysiology (e.g. neurotransmitter abnormalities). Nothing in the Wakefieldian account implies that relevant malfunctions need to be at a certain specific level, such as the level of cognitive architecture. The account is extremely flexible when it comes to the level of relevant malfunctions. They can be at any levels as long as we can talk about etiological functions at the levels.

faces which is caused by the disconnection between autonomic nervous system and face recognition system.¹⁴ This hypothesis is supported by the finding that people with Capgras delusion do not show the asymmetrical autonomic responses between familiar faces and unfamiliar faces (Ellis, Young, Quayle, & De Pauw, 1997). The experience of Capgras patient is abnormal in that it lacks the affective component that is usually a part of the experience. A Capgras delusion is formed as an explanation of this abnormal perceptual-affective experience.

Second, there might be some etiological malfunctions in attention mechanisms. According to prediction-error theories (Corlett et al., 2010; Fletcher & Frith, 2009), people with delusion have problems in the allocation of attention caused by aberrant prediction-error signals. Prediction-error signals are the indicators of the mismatch between expectations and actual inputs. They guide the process of attention allocation in such a way that attention is paid to the things or events that defy expectations. Due to aberrant prediction-error signals, people with delusion pay attention to the things or events that are not actually important. A delusion arises as the explanation of the apparent significance of these things and events. The study by Corlett et al. (2007) supports this hypothesis. In the study, two groups of participants (people with and without delusion) were tested with a task involving learning the association between certain foods and allergic reactions, while the activity of the right prefrontal cortex (which was identified as a reliable marker of prediction-error processing in previous studies) was monitored with fMRI. In the delusional group but not in the control group, the magnitude of the activity of the right prefrontal cortex was not significantly different between the cases where expectations about allergic reaction were confirmed and the cases where they were violated. In other words, in the delusional group, the right prefrontal cortex fails to distinguish prediction-error from prediction-confirmation.¹⁵

Third, the two-factor theories (Coltheart, 2007; Davies, Coltheart, Langdon, & Breen, 2001) posit the so-called “second factor”, namely, the factor that explains the adoption and/or maintenance of delusional hypotheses. The second factor is often associated with the damage to right frontal lobe. If the second factor really exists, then there can be some etiological malfunctions that are responsible for it. The main reason for positing the second factor is that it explains the difference between people with delusions and people without delusions who are experientially equivalent. For example, patients with the damage to the ventromedial prefrontal cortex do not adopt imposter hypotheses but are often regarded as experientially equivalent to people with a Capgras delusion. Unfortunately, there is no agreement on the nature of the second factor. Here is a recent proposal. Coltheart et al. (2010) argue that the second factor is the bias of neglecting contradictory evidence after adopting delusional beliefs.¹⁶ After a Capgras patient adopts the imposter hypothesis, he will be exposed to contradictory evidence such as the behavior of the “imposter”, the testimony of trustworthy friends, and so on. The contradictory evidence however is neglected because of the bias and, consequently, the delusional hypothesis is maintained.^{17,18}

4. Objections

4.1. Delusions might not be malfunctional

There are at least two kinds of possible objections to the claim that delusions are pathological because of X. First there might be some objections according to which it is not the case that all delusions are X.¹⁹ For example, an objection to the view that delusions are pathological because of their strangeness is that it is not the case that all delusions are strange. Delusional jealousy is not. Second, there might be some objections according to which X is not sufficient for a belief to be pathological. For example, an objection to the view that delusions are pathological because of the resistance to folk psychological explanations is that resisting folk psychological explanations is not sufficient for a belief to be pathological. Twisted self-deceptive beliefs resist folk psychological explanations, but they are not pathological.

In this section, I examine these two kinds of objections to my own proposal. According to the first group of objections (4.1), it is not the case that all delusions are malfunctional. Some of them are perfectly functional. According to the second (4.2), a harmful malfunction is not sufficient for a belief to be pathological. There are some beliefs that are harmful and malfunctional but are not pathological.

¹⁴ Strictly speaking, the hypothesis posits malfunctioning “connections” rather than malfunctioning “systems” or “mechanisms.” (Connected systems, such as autonomic nervous system, also behave abnormally due to the disconnection, but these abnormal behaviors are, strictly speaking, the instances of what I call “misfunction” in Section 4.2). I assume that we can talk about proper functioning or malfunctioning not only of mechanisms or systems but also of the connections between them. The assumption would be reasonable because, after all, natural selection designed not only individual mechanisms or systems but also the connections between them.

¹⁵ See (Griffiths, Langdon, Le Pelley, & Coltheart, 2014) for a methodical criticism of the study.

¹⁶ See (McKay, 2012) for critical discussions of the proposal.

¹⁷ Many two-factor theorists seem to assume something like a “central processor” that tests belief for consistency and empirical adequacy. Some might be skeptical about the assumption. But, the core ideas of two-factor theory are independent from the central processor model of belief evaluation. It is certainly conceivable that belief evaluation is achieved not by the central processor but by many different sorts of mechanisms with different functions. In that case, two-factor theorists might argue that some of those mechanisms, not the central processor, are jointly responsible for the second factor.

¹⁸ One might think, however, that those empirical studies and theories cannot support the malfunctioning thesis because they somehow already presuppose that the target mental states are malfunctional. But, this is probably not the case. The study by Corlett and colleagues, for example, simply compares people with delusions with people in control groups with fMRI during the allergy detection task, where delusions are diagnosed on the basis of psychiatric assessment criteria, not on the basis of the presence of malfunctions in the relevant sense (i.e. the failure of performing an etiological function).

¹⁹ More precisely, the objection is that it is not the case that all *pathological* delusions are X. See the footnote 1.

4.2. Psychological defence

One might think that some delusions are not malfunctional but rather are successfully performing a function, namely, the psychological defence function. In the case of B.X. that I mentioned earlier, for instance, his delusion about the fidelity of N. plays the psychological defence function. It defends B.X. from the stark reality that his body is paralyzed and N. does not love him anymore. Indeed, the idea that some delusions have defensive roles is popular recently (Bentall & Kaney, 1996; McKay, Langdon, & Coltheart, 2007). For instance, it has been suggested that persecutory delusions are produced by the so-called “externalizing attribution bias,” which is the bias of attributing negative events to other agents rather than themselves in order to defend self-esteem.

Psychological defence objection, however, is not very persuasive. I do not rule out the idea that some delusions play psychological defence roles. The problem of this objection is rather that even if they play such roles, it does not imply that they are successfully playing *biological, etiological functions*. Certainly, defending self-esteem is a good thing for us. It brings psychological comfort. But it is not obvious that it is not only psychologically good, but also biologically good. In other words, it is not obvious that defending self-esteem does not just bring psychological comfort, but also brings reproductive success. Stich famously argues that, “natural selection does not care about truth; it cares only about reproductive success” (Stich, 1990, p. 62). Similarly, we can also say that natural selection does not care about psychological comfort; it cares only about reproductive success.

Etiological functions and psychological comfort do come apart in some cases. For example, the negative emotions, such as fear or anxiety, are psychologically negative, but they play important biological functions, such as the function of avoiding dangers or threats. Presumably, we can even say that they play those functions exactly because they are psychologically negative, in the same way that pain plays the function of defending body from damage exactly because it is psychologically negative. (Pain would fail to defend body if it were psychological positive.) Furthermore, there are some conditions that are psychologically positive, but etiologically malfunctional. Nesse (1998) argues that insufficient anxiety, which is psychologically positive, is etiologically malfunctional. Anxiety has important etiological functions, and insufficient anxiety is the failure of performing these functions. People with insufficient anxiety never visit psychiatric clinics. They do not think they have to do so. Nonetheless, their anxiety mechanisms are etiologically malfunctioning and, presumably, they should be medically treated in some cases.

4.3. Doxastic shear pin

McKay and Dennett (2009) consider an interesting hypothesis according to which delusions are “doxastic shear pins.” A shear pin is a metal pin installed in complex mechanistic systems, and it is designed to break in certain circumstances in order to protect other, more expensive parts of the systems. It is conceivable that some delusions play similar roles. A possible hypothesis is that there is a mechanism whose function is to prevent motivational factors from influencing belief forming processes. But, in the situation where one faces extreme psychological stress, the mechanism is designed to break and let motivational factors influence belief forming processes in order to protect more important cognitive mechanisms. For example, in the case of B.X., the mechanism is broken, in accordance with the design, in the face of the extreme psychological stress and, consequently, his desire for the continuing fidelity of N. has a significant impact on belief forming processes, which leads to his delusion.

Mishara and Corlett (2009) propose another version of shear pin hypothesis on the basis of the prediction-error theory. According to the theory, a delusion is formed in response to prediction-error signalling abnormalities. Due to aberrant prediction-error signals, trivial things or events become abnormally salient and attention-grabbing. This is the stage of the so-called “delusional mood.” A delusion in the end arises as the explanation of the abnormal salience attached to the things and events. A delusion, according to Mishara and Corlett, can be understood as a kind of doxastic shear-pin;

The delusions appear as an *Aha-Erlebnis*, or “revelation”, concerning what had been perplexing during delusional mood. [...] The delusions are not primarily a defensive reaction to protect the self, but involve a “reorganization” of the patient’s experience to maintain behavioral interaction with the environment despite the underlying disruption to perceptual binding processes. At the Aha-moment, the “shear-pin” breaks, or as Conrad puts it, the patient is unable to shift “reference-frame” to consider the experience from another perspective. The delusion disables flexible, controlled conscious processing from continuing to monitor the mounting distress of the wanton prediction error during delusional mood and thus deters cascading toxicity. At the same time, automatic habitual responses are preserved, possibly even enhanced (Mishara & Corlett, 2009, p. 531).

Now, I do not have *a priori* reasons to rule out these hypotheses. What is crucial for me is that they are perfectly compatible with my proposal. They are compatible because there might still be some etiological malfunctions that are directly or indirectly related to delusional beliefs in those hypotheses. This is very likely in the hypothesis by Mishara and Corlett. In the hypothesis, it is assumed that prediction-error signalling is abnormal and it causes abnormalities in attention allocation processes. The role of delusions is to help people to maintain behavioral interactions with the environment despite these abnormalities. They do not eliminate the abnormalities. In the hypothesis by McKay and Dennett, the mechanism that normally constrains the influence of motivational factors on belief formation fails to perform one of its functions. It is broken. Certainly, it successfully plays another function, namely, the function of defending more important mechanisms.

Presumably, the mechanism is best understood as having two incompatible functions. On one hand, it has the function of constraining the influence of motivational factors on belief formation processes. On the other hand, it has the function of defending more important cognitive mechanisms. Those functions are incompatible because the mechanism successfully performs the latter function only by failing to perform the former and *vice versa*. In the case where the “shear pin” breaks, the mechanism is functional in the sense that it successfully performs the latter function, but it is malfunctional in the sense that it fails to perform the former.

4.4. The error management theory

The error management theory (Haselton & Buss, 2000) is the view recurrent asymmetries in the costs of false alarms shaped varieties of cognitive and behavioral biases over evolutionary history. For example, it is well-established that, compared with women, men have the stronger tendency to overperceive sexual interest. According to the error management theory, this tendency is explained by the recurrent asymmetry in the costs of errors. Smoke detectors are designed to be activated more often than they really need to be. This is because false positives are not very costly (i.e. some unnecessary evacuations), while false negatives are extremely costly (i.e. the building will be burnt down). Similarly, natural selection designed the men's sexual perception system so that it is activated more often than it really needs to be. This is because “false alarms typically result in trivial expenditures of wastes courtship effort for men: Although rejected men may experience social embarrassment, women generally do not respond antagonistically to men's overperception of sexual interest. The costs of missed mating opportunities, on the other hand, were substantial for men over the course of human evolution, because men's reproductive success can be directly affected by the access to fertile mates” (Perilloux, Easton, & Buss, 2012, p. 146).

The error management theory might cause some problems for my proposal. Maybe, some delusions are produced by the biases that evolved due to recurrent cost asymmetries. Those biases are not malfunctional. Rather, they come from the very design of relevant mechanisms. Delusional jealousy is a good candidate. False positives about the infidelity of partners do not seem to be very costly (i.e. partners will be annoyed), while false negatives are very costly (i.e. the partners might leave the subjects). So, the error management theory predicts that that jealousy is, by design, activated more often than it is really needs to be. In other words, people are designed to be oversensitive to the infidelity of partners. It is conceivable that this bias does not only explain normal jealousy but also delusional jealousy. Consistent with this hypothesis, delusional jealousy has some important characteristics in common with normal jealousy. For instance, men with delusional jealousy are especially upset about the partner's sexual infidelity, whereas women with delusion of jealousy are especially upset about the partner's emotional infidelity, which is consistent with the pattern that is seen in normal jealousy (Easton et al., 2007). If it is true that delusional jealousy is the product of the error management theoretic bias, then there is nothing malfunctional about it. As Easton and colleagues suggested, “morbid jealousy does not meet the dysfunction criterion and therefore should not be considered a mental disorder” (Easton, Schipper, & Shackelford, 2006, p. 412).

Error management theory in itself is a very plausible view. Still, I do not think that it causes serious troubles for my account. First, delusional jealousy might be pathological in the same sense that fever as the symptom of viral infection is pathological. Fever as the symptom of viral infection itself is not malfunctional. It is rather a designed defensive response. When we regard fever as pathological, we do so in virtue of the fact that it is a symptom of viral infection which involves etiological malfunctions. In other words, fever indirectly involves etiological malfunctions even though it is perfectly functional in itself. The same thing might be true about delusional jealousy. Indeed, delusional jealousy often occurs as a symptom of the conditions that are expected to involve some etiological malfunctions. It occurs, for example, in the contexts of schizophrenia, bipolar disorder, Parkinson's disease, brain injuries, and so on. When we regard delusional jealousy as pathological in those cases, we do so presumably in virtue of the fact that it is a symptom of the conditions that involve etiological malfunctions. Delusional jealousy, thus, indirectly involves etiological malfunctions even if it is perfectly functional in itself.

Second, even if normal people have the error management theoretic bias of being oversensitive to the infidelity of partners, the bias might not be sufficient to explain delusional jealousy. A possibility is that the bias is pathologically exaggerated in people with delusional jealousy. As McKay and Dennett suggested, “the most that can presently be claimed is that delusions may be produced by extreme versions of systems that have evolved in accordance with error management principles, that is, evolved so as to exploit recurrent cost asymmetries. As extreme versions, however, there is every chance that such systems manage errors in maladaptive fashion” (McKay & Dennett, 2009, p. 502). Indeed, if delusional jealousy is just an expression of an error management theoretic bias, then we cannot explain the fact that it is often seen in the contexts of schizophrenia or brain injuries. Presumably, delusional jealousy is more similar to rheumatoid arthritis, which is the product of pathologically exaggerated immune responses, than to the fever in viral infection. Another possibility is that the designed bias is only a factor of delusional jealousy. There is the second factor, which might be the bias of neglecting contradictory evidence after adopting delusional hypotheses (Coltheart et al., 2010). Maybe, the bias explains the fact that delusional jealousy is maintained, after adopted, in the absence of supportive evidence. Indeed, the two-factor theorists tend to think that the second factor is shared in all types of delusions. If so, again, it is likely that there are some malfunctions that underlie the second factor.²⁰

²⁰ Another objection that might be addressed here is that delusions do not involve any etiological malfunctions because delusion-related mechanisms do not have any etiological functions in the first place. They are rather functionless by-products (Gould, 1991; Murphy & Woolfork, 2000). See (Buller, 2005; Wakefield, 2000) for the discussions on this issue.

4.5. Harmful malfunction is not sufficient

According to the second objection, a harmful etiological malfunction is not sufficient for a belief to be pathological. For example, it is sometimes said that the fundamental idea of teleosemantics is that misrepresentations, such as non-veridical perceptions or false beliefs, involve the failures of etiological functions.

The basic idea behind teleological theories of content is that this normative notion – and its distinction between proper functioning and malfunctioning – might somehow underwrite the normative notion of content – and its distinction between representation and misrepresentation. (Neander, 1995, p. 112).

Much of the original appeal of teleosemantics was its ability to employ teleo-functional notions of purpose in order to deal with apparently normative aspects of semantic phenomena. In particular, the biological notion of failure to perform a proper function was used to attack the problem of misrepresentation, which had caused a lot of trouble for information-based theories (Godfrey-Smith, 2006, p. 62).

If it turns out that all misrepresentations involve etiological malfunctions, then, according to my proposal, all harmful misrepresentations are pathological. All harmful non-veridical perceptions and false beliefs are pathological. But, this creates too many mental disorders! Obviously, it is not the case that all harmful misrepresentations are pathological. One might falsely believe that he is not as smart as his colleagues, and the false beliefs might have a negative impact on wellbeing (e.g. the loss of self-esteem, psychological stress, amnesia, etc.). This can happen to a perfectly healthy person (although it could lead to pathological conditions such as depression). This seems to show that my proposal is wrong. More precisely, it shows that a harmful etiological malfunction is not sufficient for a mental state to be pathological. Healthy harmful misrepresentations involve harmful etiological malfunctions, according to teleosemantics, but they are not pathological.

There are some possible responses to this objection. First, I might simply reject teleosemantics. However, I do not find this option very attractive because the incompatibility with teleosemantics, which is a popular account of mental representations among naturalist philosophers of mind, is a disadvantage of my proposal.

Second, I might argue that healthy harmful misrepresentations are not harmful enough to be pathological. For example, the healthy false belief that I am not as smart as my colleagues is not a counterexample to my claim because it is not harmful enough. This option is committed to the view that the difference between healthy false beliefs and delusions is about the harmfulness condition of HDA. Both involve some kinds of etiological malfunctions. Thus, they are equivalent in terms of the malfunction condition. On the other hand, they are different in the degree of harmfulness. Delusions are harmful enough to be pathological. Healthy false beliefs are not.

This response might or might not work. But I do not agree with the idea that delusions and healthy false beliefs are different *only* in terms of the harmfulness condition. I do believe that they are also different in the malfunction condition. Teleosemantics, properly understood, does not imply that all misrepresentations involve etiological malfunctions. It is a misunderstanding of teleosemantics that all misrepresentations involve etiological malfunctions according to the theory. Thus, even if we accept teleosemantics, we do not have to accept the view that all false beliefs involve etiological malfunctions. There are different versions of teleosemantics and they need different discussions. In the following, I will talk about two notable examples; Millikan (1984), Millikan (1989) consumer-based teleosemantics and Neander (1995, 2013) informational teleosemantics.

Let us begin with the following famous example by Dretske.

Some marine bacteria have internal magnets, magnetosomes, that function like compass needles, aligning themselves (and, as a result, the bacterium) parallel to the Earth's magnetic field. Since the magnetic lines incline downward (toward geomagnetic north) in the northern hemisphere, bacteria in the northern hemisphere, oriented by their internal magnetosomes, propel themselves toward geomagnetic north. Since these organisms are capable of living only in the absence of oxygen, and since movement toward geomagnetic north will take northern bacteria away from the oxygen-rich and therefore toxic surface water and toward the comparatively oxygen-free sediment at the bottom, it is not unreasonable to speculate, as Blakemore and Frankel do, that the function of this primitive sensory system is to indicate the whereabouts of benign (i.e. anaerobic) environments (Dretske, 1991, p. 63).

What does the state of magnetosome represent? Does it represent magnetic north or the oxygen-free sediment? Millikan thinks that it represents the oxygen-free sediment, not magnetic north. In Millikan's view, what the state of magnetosome represents is determined by what the state needs to correspond to in order for the consumer of the state to perform its function successfully in its normal way. For the successful performance of the consumer (i.e. motor mechanism), the state needs to correspond to the oxygen-free sediment. After all, what is crucial for the successful functioning of the motor mechanism is to lead the bacteria to the oxygen-free sediment.

Then, Millikan's account actually allows for misrepresentations without etiological malfunctions. Suppose that I use a bar magnet to lead a bacteria upward and, consequently, the bacteria dies because of the exposure to oxygen-rich surface water. In this case, the state of the magnetosome misrepresents without etiological malfunctions. The state misrepresents because, on the one hand, it represents the oxygen-free sediment and, on the other hand, it is tokened when the oxygen-rich surface water is there instead. It does not involve any etiological malfunctions because nothing is wrong about the bacteria itself. Rather, it is just unlucky.

Still, Millikan insists that the magnetosome in the case fails to perform its function in a certain sense.

[...] Dretske is right that the magnetosome that directs that bacterium in the wrong direction because someone holds a bar magnet overhead is not broken or malfunctioning. In that sense, it is functioning perfectly properly. But it doesn't mean that it is succeeding in performing all of its functions, any more than a perfectly functional coffeemaker is performing its function when no one has put any coffee in it. Very often things fail to perform their functions, not because they are damaged, but because the conditions they are in are not their normal operating conditions (Millikan, 2004, p. 83).

The magnetosome fails to perform its functions in the same way that a coffee maker fails to perform its function (of making coffee) when nobody puts coffee beans in it. Millikan carefully distinguishes the cases where something malfunctions or, in other words, it fails to perform its function due to the intrinsic damage from the cases where something fails to perform its function due to the environmental misfortune. For the sake of avoiding confusion, I will use the term “misfunction” for the second type of failures. The coffee maker does not malfunction but misfunctions when nobody puts coffee beans in it. The magnetosome does not malfunction but misfunctions when it is fooled by my bar magnet. What is crucial here is that it might be the case that all misrepresentations involve some etiological misfunctions in Millikan's version of teleosemantics (Millikan, 1997), but it is not the case that all misrepresentations involve etiological malfunctions.

Unlike Millikan, Neander seems to be committed to the idea that all misrepresentations involve etiological malfunctions. She discusses the example of a frog (*Rana pipiens*) that catches and eats flies. The frog, however, responds not just to flies, but to other small, dark, moving things that are not flies, such as BBs. Let us call the frog's representation of its target “R.” Neander thinks that R represents small, dark, moving things rather than flies. This means that R does not misrepresent as long as it is caused by small, dark, moving things. It does not misrepresent, for instance, when it is caused by a BB. When does R misrepresent, then? It misrepresents, according to Neander, when it is caused by something which is not a small, dark, moving thing. It misrepresents, for example, when it is caused by a snail. Neander discusses a challenge according to which this view does not allow for the possibility of misrepresentation at all. After all, it is very unlikely that R is caused by a snail. In response, she argues that R will never be caused by a snail as long as perceptual systems of the frog are healthy, but “[a] sick frog might R-token at a snail if it was dysfunctional in the right way. Damaging the frog's neurology, interfering in its embryological development, tinkering with its genes, giving it a virus, all of these could introduce malfunction and error” (Neander, 1995, p. 109). These are the cases where R misrepresents, according to Neander. Then, it looks as though all misrepresentations involve some malfunctions in this account after all.

This commitment, however, is problematic. Obviously, one can have a false belief without having any neurological or genetic abnormalities. Neander recognizes this problem. Her answer to it is that the claim that misrepresentations always involve malfunctions is true only for primitive representations in the early stages of visual processing.

Consider the case where we see a skinny cow in the dim distance and mistakenly represent it as a horse (Fodor's example). Here, we may suppose, we misrepresent without malfunctioning, and clearly the content of our perceptual representation goes beyond the physical parameters of the environmental features measured. But this sophisticated representation occurs after much visual processing has already taken place, at least, this is so on computational theories of vision. In such theories, early visual processing does not represent the cow as a horse (or as a cow) but as something which looks a certain way – as having a certain outline texture, color and so on. That is, according to conventional computational theories of perception, initially there is a representation of the physical parameters of the environment as measured by the visual system. It is much plausible that there is no misrepresentation without malfunction at this level (Neander, 1995, p. 132).

The claim that misrepresentations always involve etiological malfunctions is true only in the early stages of visual processing. It is not true for sophisticated representations such as beliefs or the perceptual representations in the later stages of visual processing.

In sum, both versions of teleosemantics, Millikan's and Neander's, are actually free from the view that all misrepresentations involve etiological malfunctions. Misrepresentations might involve etiological misfunctions in Millikan's account. But, they do not always involve etiological malfunctions. In Neander's theory, misrepresentations involve etiological malfunctions in the early stages of visual processing. But, it does not generalize to other kinds of misrepresentations.

5. Conclusions

I have argued that delusions are pathological because they are harmful and malfunctional. They have significant negative impacts on wellbeing. And, some psychological mechanisms, or the connections between them, that are directly or indirectly related to delusions fail to perform their etiological functions due to intrinsic problems.

In Section 2, I discussed some possible explanations of the pathology of delusional beliefs. The explanations are not fully satisfactory primarily because they are detached from the considerations about what, in general, makes a condition pathological. My explanation, on the other hand, is an application of a general account of disorders by Wakefield, which successfully explains various kinds of physical and mental disorders.

Two types of objections were critically examined in Section 4: (1) it is not the case that all pathological delusions are malfunctional and (2) involving harmful malfunctions are not sufficient for a belief to be pathological. The first type of objections

come from the ideas that delusions are playing psychological defence functions, that they are doxastic shear pins, and that they are produced by error-management theoretic biases. In response, I argued that those ideas are perfectly compatible with the claim that delusions involve some etiological malfunctions. The second type of objection comes from the worry that all misrepresentations involve some etiological malfunctions if teleosemantics is correct. In response, I showed that the objection is based upon a misunderstanding about teleosemantics. If we are careful enough about the distinction between malfunction and misfunction, it is not very difficult to see that notable teleosemantics theories are free from such a commitment about misrepresentations.

References

- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.
- Bentall, R. P., & Kaney, S. (1996). Abnormalities of self-representation and persecutory delusions: A test of a cognitive model of paranoia. *Psychological Medicine*, 26(6), 1231–1238.
- Bortolotti, L. (2010). *Delusions and other irrational beliefs*. Oxford: Oxford University Press.
- Bortolotti, L., Broome, M. R., & Mameli, M. (2013). Delusions and responsibility for action: Insights from the Breivik case. *Neuroethics*. <http://dx.doi.org/10.1007/s12152-013-9198-4> [Advance online publication].
- Bortolotti, L., & Mameli, M. (2012). Self-deception, delusion and the boundaries of folk psychology. *Humana Mente*, 20, 203.
- Bourget, D., & Whitehurst, L. (2004). Capgras syndrome: a review of the neurophysiological correlates and presenting clinical features in cases involving physical violence. *Canadian Journal of Psychiatry*, 14, 719–725.
- Buller, D. J. (1998). Etiological theories of function: A geographical survey. *Biology & Philosophy*, 13(4), 505–527.
- Buller, D. J. (2005). *Adapting minds: Evolutionary psychology and the persistent quest for human nature*. Cambridge, MA: The MIT Press.
- Butler, P. V. (2000). Reverse Othello syndrome subsequent to traumatic brain injury. *Psychiatry*, 63(1), 85–92.
- Chadwick, P. K. (2001). Psychotic consciousness. *International Journal of Social Psychiatry*, 47(1), 52–62.
- Coltheart, M. (2007). Cognitive neuropsychiatry and delusional belief: The 33rd Sir Frederick Bartlett lecture. *The Quarterly Journal of Experimental Psychology*, 60(8), 1041–1062.
- Coltheart, M., Menzies, P., & Sutton, J. (2010). Abductive inference and delusional belief. *Cognitive Neuropsychiatry*, 15(1–3), 261–287.
- Corlett, P., Murray, G., Honey, G., Aitken, M., Shanks, D., Robbins, T., et al (2007). Disrupted prediction-error signal in psychosis: Evidence for an associative account of delusions. *Brain*, 130(9), 2387–2400.
- Corlett, P., Taylor, J., Wang, X., Fletcher, P., & Krystal, J. (2010). Toward a neurobiology of delusions. *Progress in Neurobiology*, 92(3), 345–369.
- Davies, M., Coltheart, M., Langdon, R., & Breen, N. (2001). Monothematic delusions: Towards a two-factor account. *Philosophy, Psychiatry, & Psychology*, 8(2), 133–158.
- Dretske, F. I. (1991). *Explaining behavior: Reasons in a world of causes*. Cambridge, MA: The MIT Press.
- Easton, J. A., Schipper, L. D., & Shackelford, T. K. (2006). Why the adaptationist perspective must be considered: The example of morbid jealousy. *Behavioral and Brain Sciences*, 29(4), 411–412.
- Easton, J. A., Schipper, L. D., & Shackelford, T. K. (2007). Morbid jealousy from an evolutionary psychological perspective. *Evolution and Human Behavior*, 28(6), 399–402.
- Ellis, H. D., & Young, A. W. (1990). Accounting for delusional misidentifications. *The British Journal of Psychiatry*, 157(2), 239–248.
- Ellis, H. D., Young, A. W., Quayle, A. H., & De Pauw, K. W. (1997). Reduced autonomic responses to faces in Capgras delusion. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 264(1384), 1085–1092.
- Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: A Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, 10(1), 48–58.
- Freeman, D. (2006). Delusions in the nonclinical population. *Current Psychiatry Reports*, 8(3), 191–204.
- Friedrich, J. (1993). Primary error detection and minimization (PEDMIN) strategies in social cognition: A reinterpretation of confirmation bias phenomena. *Psychological Review*, 100(2), 298.
- Gendler, T. S. (2008). Alief and belief. *The Journal of Philosophy*, 634–663.
- Godfrey-Smith, P. (1994). A modern history theory of functions. *Noûs*, 28(3), 344–362.
- Godfrey-Smith, P. (2006). Mental representation, naturalism, and teleosemantics. In G. MacDonald & D. Papineau (Eds.), *Teleosemantics* (pp. 42–68). Oxford: Oxford University Press.
- Gould, S. J. (1991). Exaptation: A crucial tool for an evolutionary psychology. *Journal of Social Issues*, 47(3), 43–65.
- Griffiths, O., Langdon, R., Le Pelley, M. E., & Coltheart, M. (2014). Delusion and prediction error: Re-examining the behavioural evidence for disrupted error signalling in delusion formation. *Cognitive Neuropsychiatry*. <http://dx.doi.org/10.1080/13546805.2014.897601> [Advance online publication].
- Haselton, M. G., & Buss, D. M. (2000). Error management theory: A new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology*, 78(1), 81.
- Hirstein, W., & Ramachandran, V. S. (1997). Capgras syndrome: A novel probe for understanding the neural representation of the identity and familiarity of persons. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 264(1380), 437–444.
- Huq, S., Garety, P., & Hemsley, D. (1988). Probabilistic judgements in deluded and non-deluded subjects. *The Quarterly Journal of Experimental Psychology*, 40(4), 801–812.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80(4), 237.
- Maier, B. A. (1974). Delusional thinking and perceptual disorder. *Journal of Individual Psychology*, 30, 98–113.
- McKay, R. (2012). Delusional inference. *Mind & Language*, 27(3), 330–355.
- McKay, R. T., & Dennett, D. C. (2009). The evolution of misbelief. *Behavioral and Brain Sciences*, 32(6), 493–510.
- McKay, R., Langdon, R., & Coltheart, M. (2005). “Sleights of mind”: Delusions, defences, and self-deception. *Cognitive Neuropsychiatry*, 10(4), 305–326.
- McKay, R., Langdon, R., & Coltheart, M. (2007). Models of misbelief: Integrating motivational and deficit theories of delusions. *Consciousness and Cognition*, 16(4), 932–941.
- Mele, A. R. (1999). Twisted self-deception. *Philosophical Psychology*, 12(2), 117–137.
- Millikan, R. G. (1984). *Language, thought, and other biological categories: New foundations for realism*. Cambridge, MA: The MIT Press.
- Millikan, R. G. (1989). Biosemantics. *The Journal of Philosophy*, 86, 281–297.
- Millikan, R. G. (2004). *Varieties of meaning: The 2002 Jean Nicod lectures*. Cambridge, MA: The MIT Press.
- Millikan, R. G. (1997). On cognitive luck: Externalism in an evolutionary frame. In P. Machamer & M. Carrier (Eds.), *Philosophy and the sciences of mind* (pp. 207–219). Pittsburgh, PA: Pittsburgh University Press and Konstanz: Universitätsverlag Konstanz.
- Mishara, A. L., & Corlett, P. (2009). Are delusions biologically adaptive?: Salvaging the doxastic shear pin. *Behavioral and Brain Sciences*, 32(6), 530–531.
- Murphy, D. (2012). The folk epistemology of delusions. *Neuroethics*, 5(1), 19–22.
- Murphy, D. (2013). Delusions, modernist epistemology, and irrational belief. *Mind & Language*, 28(1), 113–124.
- Murphy, D., & Woolfolk, R. L. (2000). The harmful dysfunction analysis of mental disorder. *Philosophy, Psychiatry, & Psychology*, 7(4), 241–252.
- Neander, K. (1995). Misrepresenting & malfunctioning. *Philosophical Studies*, 79(2), 109–141.

- Neander, K. (2013). Toward an informational teleosemantics. In D. Ryder, J. Kingsbury, & K. Williford (Eds.), *Millikan and her critics* (pp. 21–36). Chichester: John Wiley & Sons.
- Nesse, R. M. (1998). Emotional disorders in evolutionary perspective. *British Journal of Medical Psychology*, 71(4), 397–415.
- Nozick, R. (1974). *Anarchy, state, and utopia*. New York, NY: Basic Books.
- Perilloux, C., Easton, J. A., & Buss, D. M. (2012). The misperception of sexual interest. *Psychological Science*, 23(2), 146–151.
- Stich, S. P. (1990). *The fragmentation of reason: Preface to a pragmatic theory of cognitive evaluation*. Cambridge, MA: The MIT Press.
- Stone, T., & Young, A. W. (1997). Delusions and brain injury: The philosophy and psychology of belief. *Mind & Language*, 12(3–4), 327–364.
- Tengland, P. (2001). *Mental health: A philosophical analysis*. Dordrecht: Springer.
- Wakefield, J. C. (1992a). The concept of mental disorder: On the boundary between biological facts and social values. *American Psychologist*, 47(3), 373–388.
- Wakefield, J. C. (1992b). Disorder as harmful dysfunction: A conceptual critique of DSM-III-R's definition of mental disorder. *Psychological Review*, 99(2), 232–247.
- Wakefield, J. C. (2000). Spandrels, vestigial organs, and such. *Philosophy, Psychiatry, & Psychology*, 7(4), 253–269.
- Wakefield, J. C. (2011). Darwin, functional explanation, and the philosophy of psychiatry. In P. R. Adriaens & A. De Block (Eds.), *Maladapting minds: Philosophy, psychiatry, and evolutionary theory* (pp. 43–172). Oxford: Oxford University Press.